JOURNAL OF COMBINATORIAL THEORY, Series A **68**, 486–490 (1994)

Note

Catalan Numbers Revisited

DANIEL RUBENSTEIN

Department of Mathematics, University of California, Los Angeles, California 90024

Communicated by the Managing Editors

Received July 27, 1993

This paper discusses a method of proving that the number of well-formed orderings of n open and n closed parentheses is $\binom{2n}{n}/(n+1)$. More details are provided by D. S. Rubenstein ["Catalan Numbers Revisited and Generalized Catalan Numbers," 1992].

1. THE CATALAN NUMBERS AND PAIRED PARENTHESES

We define the nth Catalan number to be $C_n = \binom{2n}{n}/(n+1)$. It is known that the number of well-formed orderings of n open and n closed parentheses is C_n [3, pp. 60, 63, 64]. Many proofs of this fact involve recurrence relations and generating functions. Other proofs use combinatorial reasoning to show that the number of well-formed orderings is $\binom{2n}{n} - \binom{2n}{n-1}$, and then verify algebraically that this difference equals $\binom{2n}{n}/(n+1)$. The following proof shows that the number of well-formed orderings is C_n by partitioning the set of all $\binom{2n}{n}$ orderings into C_n subsets of cardinality $n+1$, each containing exactly one well-formed ordering.

Formulas

DEFINITION 1. A formula is any ordering of n open and n closed parentheses for any positive integer n.

For the rest of the paper, we fix n, so each formula has $2n$ parentheses (n open, n closed). The total number of formulas is $\binom{2n}{n}$.

DEFINITION 2. Let F be a formula. We denote the ith parenthesis of F by F_i. If G is another formula (including the possibility that $G = F$) and

0097-3165/94 $6.00

F_i and G_j are both open or both closed, we write $F_i \sim G_j$. If one is open and the other closed, we write $F_i \nsim G_j$. If $\mathscr{S}$ is a subset of $\{1, 2, \ldots, 2n\}$ then $F_{\mathscr{S}} = \{F_k | k \in \mathscr{S}\}$. For simplicity we write $F_{i_1, i_2, \ldots, i_k}$ instead of $\{F_{i_1}, F_{i_2}, \ldots, F_{i_k}\}$. We use $[i, j]$ to represent $\{i, i+1, i+2, \ldots, j\}$.

ALGORITHM. *We pair parentheses in any formula F in the following manner. We first pair F_i and F_{i+1} whenever F_i is open and F_{i+1} is closed. We then proceed to pair open parentheses N_i with closed parentheses N_j by induction on the distance d between them. Suppose that all such pairings of open F_i with closed F_j $(0 < j - i < d)$ have been made. We then pair an open F_i with a closed F_{i+d} if they are not yet paired. If F_i is paired with F_j, we say that $F_{i,j}$ pairs. Parentheses which are not paired by the algorithm are called unpaired. If every parenthesis of F is a member of a pair, we say that F is well-formed.*

The pairs constructed by this algorithm do not intertwine. In other words, if $a < b < c < d$, $F_{a,c}$ and $F_{b,d}$ are not both paired.

DEFINITION 3. We say that $F_{[i,j]}$ is a *block* if $F_{i,j}$ pairs and every parenthesis F_k $(i < k < j)$ is paired with some F_l $(i < l < j)$. We make the convention that the empty set is a block as well.

Note that blocks contain the same number of open parentheses as closed, and thus are always of even length. Any parenthesis which is a member of a pair is also a member of at least one block. Any parenthesis which is unpaired cannot be a member of a block.

Our pairing algorithm works inductively on the distance between parentheses. First it forms all blocks of size 2. The next step in the algorithm forms all blocks of size 4, and so on. By induction, it is clear that the question of whether or not $F_{[i,j]}$ forms a block depends only on its members. The parentheses outside a block $F_{[i,j]}$ have no effect on the pairing within $F_{[i,j]}$. We therefore have:

LEMMA 1. *If F and G are formulas where $F_{[i,j]}$ is a block, and $G_k \sim F_k$ for all k, $(i \le k \le j)$, then $G_{[i,j]}$ is a block as well.*

DEFINITION 4. Let $\mathscr{N} = \{i_1, i_2, \ldots, i_t, j_t, j_{t-1}, \ldots, j_1\}$, where $i_1 < i_2 < \cdots < i_t < j_t < j_{t-1} < \cdots < j_1$. If $F_{i_1, j_1}, F_{i_2, j_2}, \ldots, F_{i_t, j_t}$ are pairs, we say that $F_{\mathscr{N}}$ is a nest. If $F_{i_1, i_2, \ldots, i_t}$ is a set of unpaired closed parentheses in F and $F_{j_t, j_{t-1}, \ldots, j_1}$ is a set of unpaired open parentheses in F, we say that $F_{\mathscr{N}}$ forms an *antinest*.

DEFINITION 5. An outer *nest* is a nest $F_{\mathscr{O}}$ such that every pair $F_{i,j}$ surrounding a member of the nest $F_{\mathscr{O}}$ is in $F_{\mathscr{O}}$ as well. An equivalent definition is that $F_{\mathscr{O}}$ is an outer nest if the spaces between consecutive members of $F_{\mathscr{O}}$ consist only of blocks.

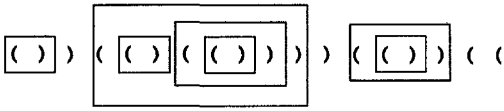

FIG. 1. An example of how one pairs a formula with $n = 11$. A rectangle is drawn around each block. Two parentheses form a pair if one is the leftmost parenthesis in a rectangle and the other is the rightmost parenthesis in the same rectangle.

LEMMA 2. *Any well-formed formula has n outer nests.*

Proof. Every nest has an innermost pair, and every pair is the innermost pair of at least one nest. If $F_{i,j}$ is the innermost pair of the outer nest $F_{\mathscr{O}}$, then $F_{\mathscr{O}}$ must contain all pairs which surround $F_{i,j}$ and can contain no other pairs. From this we see that there is one and only one outer nest with innermost pair $F_{i,j}$. Since a well-formed formula has n pairs, it has n outer nests.

Since the number of paired open parentheses equals the number of paired closed parentheses, the number of unpaired open parentheses equals the number of unpaired closed parentheses. It is clear from the pairing algorithm that all unpaired closed parentheses are to the left of all unpaired open parentheses. Thus the unpaired parentheses form an antinest. From the definition of a block, the spaces between consecutive members of this antinest consist of blocks.

Given any formula F, let G be the formula defined by $G_i \sim F_i$ if F_i is paired, and $G_i \nsim F_i$ if F_i is unpaired. Thus G is obtained from F by reversing the unpaired parentheses of F.

DEFINITION 6. Let $\mathscr{U}(F) = \{k \mid F_k \text{ is unpaired}\}$.

Note that $k \in \mathscr{U}(F)$ if and only if $G_k \nsim F_k$. Also, $F_{\mathscr{U}(F)}$ is the antinest of all unpaired parentheses.

LEMMA 3. *If G and F are defined as above, then $G_{\mathscr{U}(F)}$ is an outer nest and G is well-formed.*

Proof. Since $F_{\mathscr{U}(F)}$ is the antinest of all unpaired parentheses, any closed parentheses of $F_{\mathscr{U}(F)}$ are to the left of any open parentheses of $F_{\mathscr{U}(F)}$. By Lemma 1, blocks of F are also blocks of G. Since the members of $F_{\mathscr{U}(F)}$ are all unpaired, they do not lie within pairs. Closed parentheses of $G_{\mathscr{U}(F)}$ are to the right of open parentheses of $G_{\mathscr{U}(F)}$. Since all other parentheses pair in G as they do in F, the members of $G_{\mathscr{U}(F)}$ pair among themselves, and all must pair since there are no open parentheses to the

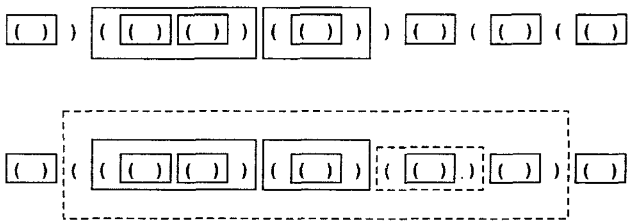

FIG. 2. A formula with $n = 11$ and the well-formed formula to which it is mapped. Note that the unpaired parentheses then pair amongst themselves in a nested fashion. Blocks are boxed as in the previous figure. The new blocks created by the flipping are dotted.

right of any closed parentheses in $G_{\mathscr{U}(F)}$. Thus $G_{\mathscr{U}(F)}$ satisfies the properties of an outer nest, and all parentheses in G pair.

DEFINITION 7. Let G be a well-formed formula and let $\mathscr{S}(G)$ be the set $\{F\}$ of formulas which are mapped to G by reversing the parentheses in $F_{\mathscr{U}(F)}$. Since each formula gets mapped to one and only one well-formed formula, we have that the set of all formulas equals the disjoint union $\bigcup_G \mathscr{S}(G)$, G well-formed.

Given a formula F and a well-formed formula G, if $F \in \mathscr{S}(G)$ then G can be obtained from F by reversing the direction of the unpaired parentheses of F. By again reversing these parentheses, we see that we can also obtain F from G.

If $F \in \mathscr{S}(G)$ then $G_{\mathscr{U}(F)}$ is an outer nest, so any $F \in \mathscr{S}(G)$ can be obtained by reversing some outer nest in G. Conversely, let $G_{\mathscr{O}}$ be an outer nest and reverse its parentheses to obtain F'. Since the spaces between members of $G_{\mathscr{O}}$ consist of blocks, all parentheses of F' not in $F'_{\mathscr{O}}$ pair among themselves in a fashion similar to those not in $G_{\mathscr{O}}$. Therefore $F'_{\mathscr{O}}$ consists of closed parentheses to the left open parentheses. Since all other parentheses are members of pairs, $F'_{\mathscr{O}}$ remain unpaired, and so $\mathscr{U}(F') = \mathscr{O}$. Thus $F' \in \mathscr{S}(G)$.

The above tells us that the cardinality of $\mathscr{S}(G)$ equals the number of outer nests in G plus one (since $G \in \mathscr{S}(G)$). Thus by Lemma 2 the cardinality of $\mathscr{S}(G)$ equals $n + 1$ for every G, and each set $\mathscr{S}(G)$ contains one and only one well-formed formula. Since there are altogether $\binom{2n}{n}$ formulas, the number of well-formed formulas is $\binom{2n}{n}/(n + 1)$.

Another algorithm which maps $\binom{2n}{n}$ orderings of parentheses (or in this case, south–east moving graphs) into groups of size $n + 1$ is given in "Counting Plane Trees," by David Feldman from the University of New Hampshire [4]. The algorithms match well-formed formulas with different sets of n formulas, and so clearly the algorithms are different.

REFERENCES

1. D. S. RUBENSTEIN, "Catalan Numbers Revisited and Generalized Catalan Numbers," technical report, MIT, Cambridge, MA, May 18, 1992.
2. R. P. STANLEY, "Enumerative Combinatorics," Vol. I, Wadsworth & Brooks/Cole, Monterey, CA, 1986.
3. D. STANTON AND D. WHITE, "Constructive Combinatorics," Springer-Verlag, New York, 1986.
4. D. FELDMAN, Counting plane trees, manuscript, University of New Hampshire. [unpublished]